Asleep at the Keyboard? Assessing the Security of GitHub Copilot’s Code Contributions

Year: 2021 | Citations: 558 | Authors: H. Pearce, Baleegh Ahmad, Benjamin Tan, Brendan Dolan-Gavitt, R. Karri

---

Abstract

There is burgeoning interest in designing AI-based systems to assist humans in designing computing systems, including tools that automatically generate computer code. The most notable of these comes in the form of the first self-described ‘AI pair programmer’, GitHub Copilot, which is a language model trained over open-source GitHub code. However, code often contains bugs—and so, given the vast quantity of unvetted code that Copilot has processed, it is certain that the language model will have learned from exploitable, buggy code. This raises concerns on the security of Copilot’s code contributions. In this work, we systematically investigate the prevalence and conditions that can cause GitHub Copilot to recommend insecure code. To perform this analysis we prompt Copilot to generate code in scenarios relevant to high-risk cybersecurity weaknesses, e.g. those from MITRE’s “Top 25” Common Weakness Enumeration (CWE) list. We explore Copilot’s performance on three distinct code generation axes—examining how it performs given diversity of weaknesses, diversity of prompts, and diversity of domains. In total, we produce 89 different scenarios for Copilot to complete, producing 1,689 programs. Of these, we found approximately 40% to be vulnerable.